Cancer Genetics 208 (2015) 408–413

Cancer Genetics

BRIEF COMMUNICATION

Concordance of copy number alterations using a common analytic pipeline for genome-wide analysis of Illumina and Affymetrix genotyping data: a report from the Children's Oncology Group

Marijana Vujkovic [a,*], Edward F. Attiyeh [a], Rhonda E. Ries [b], Michelle Horn [a], Elizabeth K. Goodman [a], Yang Ding [a], Marko Kavcic [c], Todd A. Alonzo [d], Robert B. Gerbing [e], Betsy Hirsch [f], Susana Raimondi [a], Alan S. Gamis [g], Soheil Meshinchi [b,1], Richard Aplenc [a,1]

[a] *Division of Oncology, The Children's Hospital of Philadelphia, Philadelphia, PA, USA;* [b] *Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA;* [c] *University Medical Centre, Ljubljana, Slovenia;* [d] *Keck School of Medicine, University of Southern California, Los Angeles, CA, USA;* [e] *Children's Oncology Group, Monrovia, CA, USA;* [f] *University of Minnesota, Minneapolis, MN, USA;* [g] *Division of Hematology/Oncology/Bone Marrow Transplantation, Children's Mercy Hospitals and Clinics, Kansas City, MO, USA*

Copy number alterations (CNAs) are a hallmark of pediatric cancer genomes. An increasing number of research groups use multiple platforms and software packages to detect and analyze CNAs. However, different platforms have experimental and analysis-specific biases that may yield different results. We sought to estimate the concordance of CNAs in children with de novo acute myeloid leukemia between two experimental platforms: Affymetrix SNP 6.0 array and Illumina OmniQuad 2.5 BeadChip. Forty-five paired tumor-remission samples were genotyped on both platforms, and CNAs were estimated from total signal intensity and allelic contrast values using the allele-specific copy number analysis of tumors (ASCAT) algorithm. The two platforms were comparable in detection of CNAs, each missing only two segments from a total of 42 CNAs (4.6%). Overall, there was an interplatform agreement of 96% for allele-specific tumor profiles. However, poor quality samples with low signal/noise ratios showed a high rate of false-positive segments independent of the genotyping platform. These results demonstrate that a common analytic pipeline can be utilized for SNP array data from these two platforms. The customized programming template for the preprocessing, data integration, and analysis is publicly available at https://github.com/AplenCHOP/affyLumCNA.

Keywords Pediatrics, acute myeloid leukemia, SNP array, copy number alterations
Published by Elsevier Inc.

Changes in DNA copy number occur frequently in tumor genomes (1–3) and can be measured with single nucleotide polymorphism (SNP) array technologies, such as those from Illumina (San Diego, CA) and Affymetrix (Santa Clara, CA). These platforms differ in experimental methodology, with the Affymetrix SNP array using oligonucleotide hybridization on fixed glass slides, whereas Illumina's technology employs *in situ* single nucleotide extension reactions on bead arrays (4). The comparability of output from these two arrays has been reviewed regarding data quality, copy number alteration (CNA) calling, breakpoint accuracy, reproducibility, and concordance (4,5). These reviews reported an interplatform concordance of less than 50%. Data analysis is additionally complicated by technology-specific output formats (6). Gunnarsson *et al.* described that tools for detecting CNAs from Illumina and Affymetrix raw data from identical samples typically do not yield identical CNA calls and insufficiently

Received November 12, 2014; received in revised form April 27, 2015; accepted April 28, 2015.

* Corresponding author.
E-mail address: vujkovicm@email.chop.edu

[1] Both authors contributed equally to this work.

2210-7762/$ - see front matter Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.cancergen.2015.04.010

discriminate CNA artifacts from true CNA calls (7). Furthermore, CNA-detection tools for Illumina and Affymetrix often yielded different CNA calls, with Affymetrix detecting a higher degree of CNAs and Illumina detecting higher rates of loss of heterozygosity (LOH) (4,5). However, as SNP analytic packages improve, data variability may be overcome using appropriate data preprocessing (6). Currently numerous tools exist for array analysis of cancer samples on multiple platforms (e.g., allele-specific copy number analysis of tumors [ASCAT], OncoSNP, generalized pair hidden Markov model, Genome Alteration Print, and PennCNV-Tumor) (8–10). We aimed to integrate and analyze CNA data from the Affymetrix and Illumina platforms and test the concordance between these two platforms in a set of matched tumor-remission DNA samples from pediatric acute myeloid leukemia (AML) patients genotyped on the Affymetrix SNP 6.0 array and Illumina OmniQuad 2.5 BeadChip.

Materials and methods

A total of 90 matched tumor-remission samples from 45 children with newly diagnosed AML were obtained from the Children's Oncology Group trials AAML0531 (11), AAML03P1 (12), and CCG-2961 (13). Cytogenetic abnormalities were determined by standard chromosome-banding analysis. Paired tumor-remission samples were genotyped on the Affymetrix SNP 6.0 chip at the Hudson Alpha Institute, Birmingham, AL, and on the Illumina 2.5M OmniQuad chip at the Children's Hospital of Philadelphia, Philadelphia, PA. The Affymetrix 6.0 SNP array contains 946,000 polymorphic and nonpolymorphic copy number probes with a 2.18-KB median intermarking spacing. The Illumina 2.5M BeadChip contains 2,390,000 SNPs with a median 0.64-KB probe spacing. The BeadChip does not contain any nonpolymorphic CNV loci by design. To calculate log R ratio (LRR) and B-allele frequency (BAF) values, raw Affymetrix CEL files were converted to allele-specific signals and genotype calls using Affymetrix Power Tools and the BirdSeed algorithm. All markers were updated to hg19, GRCh37 using LiftOver. Total signal intensity (LRR) was obtained by summing allele-specific intensities and was normalized with HapMap3 data. BAF, or allelic contrast, was calculated as the normalized ratio of the quantity of the variant allele to the total quantity of both alleles. For the Illumina samples, BAFs and LRRs were exported directly from Illumina GenomeStudio and signal intensity values corrected for patterns of genomic wave using PennCNV (14).

A matched (tumor vs. remission) ASCAT was performed using ASCAT 2.2 in R (15). ASCAT was run with default parameters. All copy number profiles were manually reviewed (BG or MV). For the manual review, the CNA segments identified *in silico* by the ASCAT algorithm were reviewed by examining the LRR and BAF visualizations. Illumina GenomeStudio GenomeViewer and Partek Genomics Suite (St. Louis, MO) for Affymetrix data were used as necessary. In a subsequent manual review, one study member (MV) was blinded to the ASCAT copy number output and then visually identified CNAs from LRR and BAF plots. False-negatives were manually annotated, and the final list of manually reviewed CNA calls was defined as the gold standard. Platform-specific sensitivity and specificity were estimated by comparing the number of CNAs from the gold standard with the ASCAT output generated for Illumina and Affymetrix data. All altered segments were classified as CNA (+), and normal diploid segments were classified as CNA (−). Sensitivity was defined as the number of CNAs identified *in silico* divided by the total number of true CNAs as identified by visual inspection. Specificity was defined as the number of samples with no computationally identified CNAs divided by the total number of samples in which no CNA was visually detected. The interobserver Cohen's κ statistic was used to test the platform agreement in CNA calls.

Results

Of 45 patients, 43 were successfully genotyped on both platforms. Two samples failed quality control and were excluded from further analysis. Table 1 summarizes the observed CNAs. We observed 42 CNAs, consisting of 18 losses, six regions of copy-neutral (CN) LOH, and 18 gains. In three patients with a normal karyotype, the SNP array detected an additional region of CN-LOH. Eight patients had gain of chromosome 8, which was observed with both cytogenetic analysis and genotyping, except for one patient whose cytogenetic profile was unknown. In four patients, the gain of chromosome 8 was the sole abnormality identified by G-banding, but the SNP array detected an additional CNA in two: a deletion of 7q and 9q. Other cytogenetically identified chromosomal trisomies and monosomies were detected with SNP arrays as well. One CN-LOH on chromosome 13 was visually identified based on BAF values but was not detected with the Affymetrix or Illumina array, because of low allelic contrast. Cytogenetic analysis detected nine balanced inversions or translocations, which, as expected, were not detected by SNP arrays. However, in two of these cases, both involving inversion 16 (*CBFB-MYH11* fusion), the SNP array detected two cryptic abnormalities: an associated deletion of the *MYH11* gene in one and a secondary deletion of 7q36.1-7q36.2 in the other. In total, the SNP array identified four chromosomal losses, not visible by G-banding, and seven regions of CN-LOH that were not detectable by cytogenetic analysis. In contrast, cytogenetic analysis identified one gained region—specifically, add(2)(q35) in patient 11—that was not detected by array. The arrays further characterized some of the cytogenetically visible abnormalities. For example, in patient 14, the array showed that the additional material translocated to 7q31—the add(7)(q31)—was an extra copy of the distal long arm of chromosome 1. Of the remaining 23 patients without observable CNAs, seven had a normal karyotype and others had either inversions or translocations or both. Two patients had an unknown karyotype with no insertions or deletions.

Using the manually reviewed CNAs as the gold standard, ASCAT correctly called the majority of true CNAs, independent of the platform. The smallest CNA detected was a 358-KB loss on chromosome 10 and was detected on both platforms. Overall, the Illumina platform had a sensitivity of 95.3% (95% confidence interval [CI] 82.9–99.2%) with ASCAT and a specificity of 33.6% (95% CI 31.9–35.3%). The Affymetrix platform had a similar sensitivity, 95.3% (95% CI 82.9–99.2%), and a specificity of 33.2% (95% CI 31.6–34.9%). The interplatform agreement was high, with a